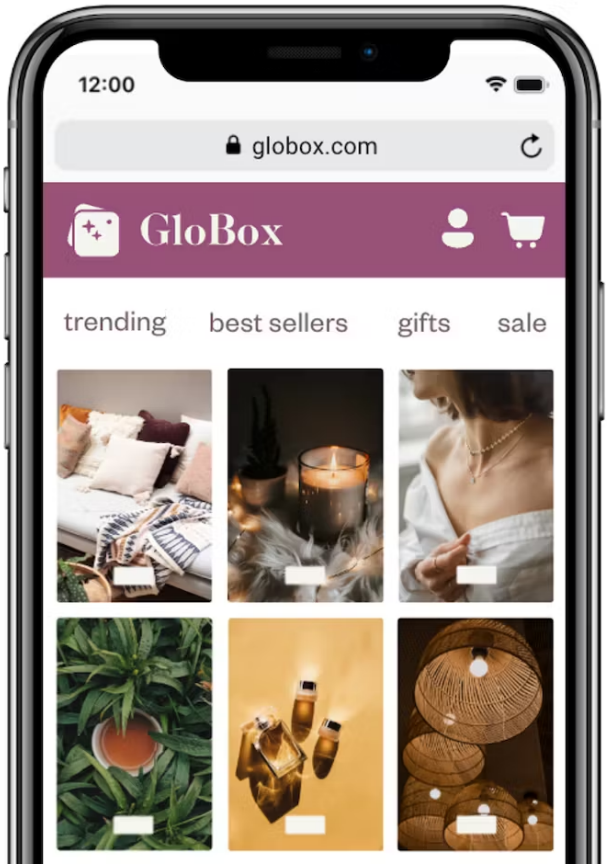
Globox A/B Test Analysis Report

Date: 3rd August 2023
Completed By: Ben Campbell, Data Analyst, Globox

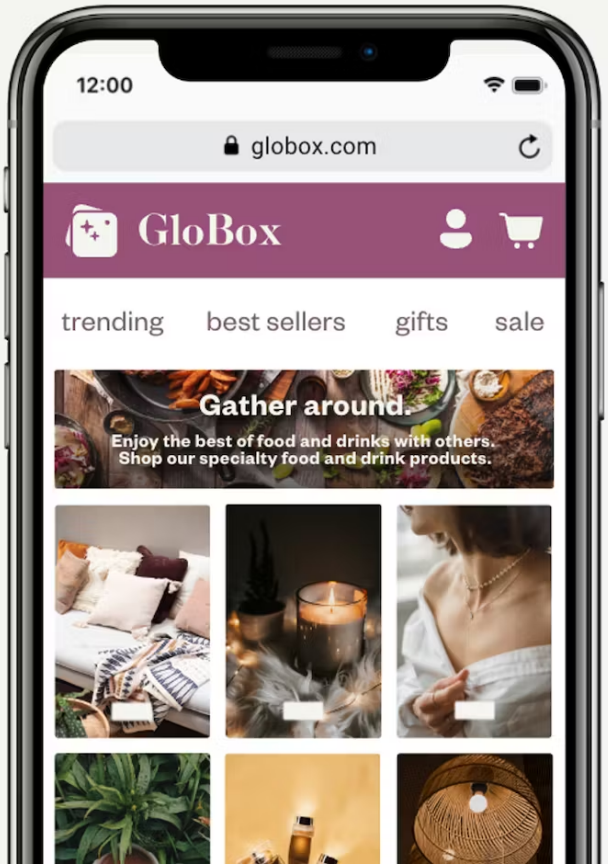
Group A: Control

existing landing page



Group B: Treatment

landing page with food & drink banner



Executive Summary

Note key focus areas highlighted in **Bold.**

An A/B Test was conducted between 25th January 2023 - 6th February 2023 (13 days), to measure conversion rates and average spend with **a total of 48,943 users**, **both conversion rate (+0.71%) and average spend (+$0.02) reported increases**.

Additional findings:

- **Female Conversion Rate increased by 0.29% but Average Spend decreased by $0.33**
- **Male Conversion Rate increased by 1.17% and Average Spend increased by $0.35**
- Conversion rates on devices all increased by a minimum of 0.6%, but average spend decreased in both iOS and Not Disclosed with Android increasing by $0.16
- 10 out of 12 countries report an increase in conversion rate between the treatment and control group
- 6 out of 12 countries report an increase in average spend with a net increase of $0.02, with the UK having the highest increase $2.39

Conclusion & Recommendations:

The A/B test was conducted to determine the effectiveness of a new feature on Globox's platform, using two metrics: conversion rate and average amount spent per user.

We are 95% confident that the true difference between the Conversion Rate is between 0.0035 and 0.0107 and we are 95% confident that the true difference between the Average Spend is between -0.439 and 0.471 (Confidence Intervals).

The results showed that there **was a statistically significant difference in the conversion rate between the control and treatment groups**, with the treatment group showing a higher conversion rate. However, **there was no statistically significant difference in the average amount spent per user between the two groups**.

Based on these findings, we recommend that Globox should not launch the new feature to all users even though it has the potential to increase conversion rates since further analysis to understand the impact of the new feature on revenue is recommended. Based on this additional analysis, stakeholders can make an informed decision about whether to launch the experiences to all users.

Recommendations are to adjust parameters (duration of AB test, increase sample size), define success and ROI, focus on demographics which reported an increase in spend and

conversion rates. Based on the **Power Analysis** carried out an extension of the sample size would be required and there is no clear evidence to suggest the presence of a **novelty effect**.

Scope

Project Background

GloBox prides itself on offering a curated selection of products that evoke a sense of adventure and bring the world to customers' doorsteps. The company's existing customer base is well-acquainted with its boutique fashion items and high-end decor products. However, the recent expansion of the food and drink category presents an opportunity to further enhance revenue streams. To ensure effective promotion and increase awareness of this product category, the Growth team has devised an A/B test strategy.

The A/B test will involve two groups: the control group and the test group. The control group will experience the standard browsing experience on the GloBox website, without any banner highlighting the food and drink category. On the other hand, the test group will be exposed to a prominently displayed banner featuring key food and drink products at the top of the website (see Figure 1).

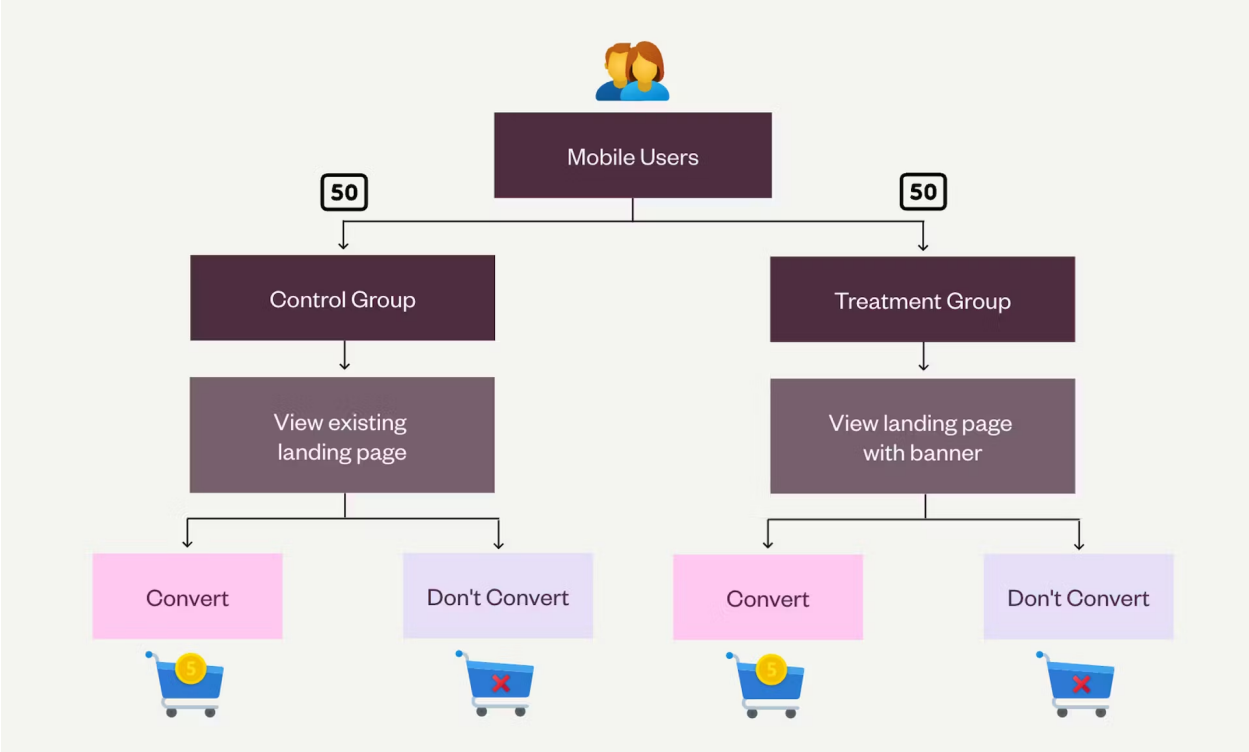


Figure 1

Project Process

The project process was as follows:

1. Data Extraction:
    a. Extracted data from the relational database into a csv file to enable in-depth analysis.
    b. Utilised SQL queries to extract the user-level aggregated dataset and consolidate relevant tables for analysis.
2. Data Analysis:
    a. Employed SQL to bring together the extracted data tables and conduct thorough analysis.
    b. Utilised inferential statistics, including confidence intervals and hypothesis tests, to identify areas of concern and opportunities for learning from successes.
3. Data Visualization:
    a. Expanded on the initial analysis by leveraging spreadsheets to design visualisations in Tableau.
    b. Created visual representations to identify and highlight trends within the dataset effectively.
4. Recommendations and Conclusion:
    a. Provided recommendations and conclusions based on the analysis findings.

Note that the data analysis team did not set the test parameters, and the recommendations were made accordingly. Please refer to the Recommendations section of the report for more details.

The Business Goals

Inferential statistics play a crucial role in determining the significance and meaningfulness of the changes observed through A/B testing. For GloBox, the focus lies on examining two key business goals: the user conversion rate and the average amount spent per user. By conducting rigorous statistical analysis, we aim to assess whether the introduction of the new food and drink banner is leading to noticeable changes in these metrics.

Success for GloBox in relation to the A/B test would be characterised by the following:

1. **Increased User Conversion Rate:**

The primary goal is to observe a statistically significant and meaningful increase in the user conversion rate for the food and drink category. This would indicate that the introduction of the banner has effectively captured user attention and motivated them to make purchases within this category. The test group should exhibit a significantly higher conversion rate compared to the control group.

2. **Improved Average Spend per User:**

Another crucial objective is to determine whether the new banner influences the average amount spent per user within the food and drink category. Success would be evident if the test group exhibits a statistically significant and meaningful increase in average spend compared to the control group. This would indicate that the banner has not only attracted users but has also prompted them to spend more on food and drink products.

To assess these business goals, appropriate statistical tests, such as hypothesis testing or confidence interval estimation, will be employed. The results will provide insights into the effectiveness of the food and drink banner in driving user engagement, conversion rates, and overall revenue. By achieving significant improvements in user conversion rate and average spend, GloBox can consider the A/B test a success, signifying the banner's efficacy in enhancing customer engagement and driving revenue growth within the food and drink category.

Key Findings (Inferential Statistics)

The AB test was conducted during 25th January 2023 through to 6th January 2023 with a total sample size of 48,493

2.1:Key Findings 25th January 2023 & 6th February 2023:

Metric	Group A: Control	Group B: Treatment	Grand Total	P Value
Total Number of Users	**24,343**	**24,600**	**48,943**	**N/A**
Conversion Rate	**3.92%**	**4.63%**	**4.28%**	**0.0001**
Average Amount Spend $	**$3.37**	**$3.39**	**$3.38**	**0.9438**

The data provided shows that the treatment group (Group B) has a higher conversion rate (4.63%) compared to the control group (Group A) with a conversion rate of 3.92%. However, the average amount spent by users in both groups is quite similar, with Group A having an average of $3.37 and Group B with $3.39, resulting in a negligible difference of $0.02.

Further analysis and calculation were conducted in the form of Hypothesis Testing and Confidence Intervals to determine the statistical significance of the A/B Test Data.

2.2: Hypothesis Testing Results

Hypothesis test 1 - Conversion Rate

Null and alternative hypotheses:
Null hypothesis (H0): The conversion rate is the same between Group A (control group) and Group B (treatment group).
Null hypothesis (H0): $\mu A = \mu B$

Alternative hypothesis (Ha): There is a difference in the conversion rate between Group A and Group B.
Alternative hypothesis (Ha): $\mu A \neq \mu B$

Significance level= 0.05/5%

See spreadsheet Globox A/B Test GoogleSheets Analysis for workings and tab Conversion Rate Hypothesis

Conclusion
p = 0.0001, statistically significant. **We reject the null hypothesis that there is no difference in the user conversion rate between the control and treatment.**

Hypothesis test 2 - Average Spend
Null hypothesis (H0): There is no difference in the average amount spent per user between the Control and Treatment groups.
Null hypothesis (H0): $\mu 1 = \mu 2$

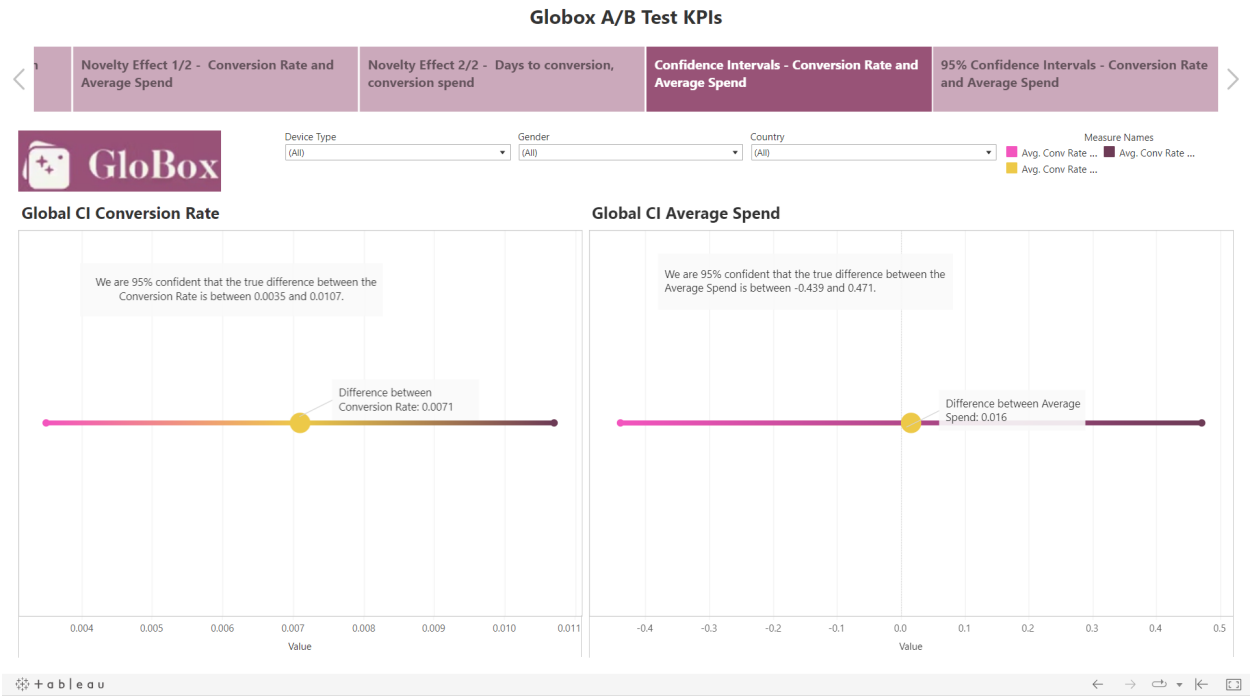
Alternative hypothesis (HA): There is a difference in the average amount spent per user between the Control and Treatment groups.
Alternative Hypothesis ($(\mu 1 \neq \mu 2)$

Significance level= 0.05/5%

See spreadsheet Globox A/B Test GoogleSheets Analysis for workings and tab Avg. Spend Hypothesis

Conclusion
p = 0.944, statistically insignificant. **We fail to reject the null hypothesis that there is no difference in the mean amount spent per user between the control and treatment.**

2.3: Confidence intervals



Metric	Group Mean	Lower Bound	Upper Bound
Average Amount Spent	$0.016	$-0.0439	$0.0471
Conversion Rate	0.71%	0.35%	1.07%

Average Amount Spent:
For the Control Group, the average amount spent is $3.37 and the Treatment Group average amount spent is $3.39, with a confidence interval ranging from -$0.0439 to $0.0471. The 95% confidence interval means that if we were to repeat the experiment 100 times, we would expect the true difference to be within -$0.0439 to $0.0471.

We can conclude that there is no statistically significant difference in the average amount spent between Control Group and Treatment Group. In other words, the data does not provide sufficient evidence to suggest that the treatment (Group B) had a significant impact on the average amount spent compared to the control group (Group A).

Conversion Rate:
For the Control Group, the conversion rate is 3.92% and for the Treatment Group the conversion rate is 4.63%, with a confidence interval ranging from 0.35% to 1.07%. The 95% confidence interval means that if we were to repeat the experiment 100 times, we would expect the true difference to be within 0.35% to 0.71%.

We can conclude that there is a statistically significant difference in the conversion rates between Control Group and Treatment Group. In other words, the data does provide sufficient evidence to suggest that the treatment (Group B) had a significant impact on the conversion rate compared to the control group (Group A).

See Tableau Globox AB Test Analysis story point 95% Confidence Intervals - Conversion Rate and Average Spend

Supplementary Findings

Table 2.5: Gender Metrics

Distribution

Gender	Group A: Control	Group B: Treatment	Grand Total
Female	10,069	10,061	20,130
Male	10,054	10,235	20,289
Not Disclosed	3,412	3,443	6,855
Other	808	861	1,669
Grand Total	**24,343**	**24,600**	**48,943**

Conversion Rate

Gender	Group A: Control	Group B: Treatment	Grand Total	Differential B-A (+/-)
Female	5.14%	5.44%	5.29%	0.29%
Male	2.63%	3.79%	3.21%	1.17%
Not Disclosed	4.31%	5.17%	4.74%	0.86%
Other	3.22%	3.02%	3.12%	-0.20%
Grand Total	**3.92%**	**4.63%**	**4.28%**	**0.71%**

Average Spend

Gender	Group A: Control	Group B: Treatment	Grand Total	Differential B-A (+/-)

Female	$4.46	$4.13	$4.30	-$0.33
Male	$2.25	$2.60	$2.43	$0.35
Not Disclosed	$3.62	$3.74	$3.68	$0.11
Other	$2.77	$2.77	$2.77	$0.00
Grand Total	**$3.37**	**$3.39**	**$3.38**	**$0.02**

The data reveals the gender distribution across two groups: Control (Group A) and Treatment (Group B). Females make up approximately 51.6% of the total participants, with 10,069 females in the Control group (49.9% of the group) and 10,061 females in the Treatment group (51.5% of the group). Males represent around 48.4% of the total participants, with 10,054 males in the Control group (49.7% of the group) and 10,235 males in the Treatment group (52.5% of the group). Additionally, the "Not Disclosed" category accounts for about 14% of the total participants, and the "Other" category comprises roughly 3.4%.

The provided data includes conversion rate percentages and average spend for different genders across two groups: Control (Group A) and Treatment (Group B). **The analysis shows that the female conversion rate increased by 0.29% in the Treatment group compared to the Control group, but the average spend decreased by $0.33. On the other hand, the male conversion rate saw a significant increase of 1.17% in the Treatment group, accompanied by an average spend increase of $0.35.** These observations suggest that the Treatment group's strategies had a more positive impact on male conversion rates and spending compared to females, although the average spend for both genders decreased slightly in the Treatment group.

Table 2.6: Device Metrics

Distribution

Device Type	Group A: Control	Group B: Treatment	Grand Total
Android	15,054	15,235	30,289
iOS	9,142	9,218	18,360
Other	147	147	294
Grand Total	**24,343**	**24,600**	**48,943**

Device Type	Group A: Control	Group B: Treatment	Grand Total	Differential B-A (+/-)
Android	2.77%	3.52%	3.15%	0.75%

iOS	5.85%	6.47%	6.16%	0.62%
Other	2.04%	4.08%	3.06%	2.04%
Grand Total	**3.92%**	**4.63%**	**4.28%**	**0.71%**

Device Type	Group A: Control	Group B: Treatment	Grand Total	Differential B-A (+/-)
Android	$2.31	$2.47	$2.39	$0.16
iOS	$5.07	$4.92	$5.00	-$0.15
Other	$6.54	$3.46	$5.00	-$3.08
Grand Total	**$3.37**	**$3.39**	**$3.38**	**$0.02**

The provided data presents the distribution of device types across two groups: Control (Group A) and Treatment (Group B). The total number of Android users was significantly higher than iOS users, with Android representing approximately 61.9% of the total participants and iOS representing about 37.5%.

The data shows that conversion rates increased across all device types in the Treatment group, with a minimum increase of 0.6%. However, the average spend decreased for iOS and the "Not Disclosed" category, while it increased by $0.16 for Android users. These observations indicate that the Treatment group's strategies had a positive impact on conversion rates across all devices, but the effect on average spend varied depending on the device type, with Android users experiencing a slight increase and iOS and "Not Disclosed" users experiencing a decrease.

Table 2.7: Country Conversion Rates (Ascending Order):

Country	Number of users	Group A - Control	Group B - Treatment	Conversion Rate	Differential B-A (+/-)
Canada	**1,570**	4.69%	6.48%	5.61%	1.79%
Mexico	**5,738**	2.95%	4.45%	3.71%	1.50%
Germany	**3,854**	3.20%	4.41%	3.81%	1.21%
France	**3,090**	3.13%	4.18%	3.66%	1.05%
Australia	**1,168**	2.14%	3.04%	2.57%	0.90%
UK	**2,949**	2.89%	3.68%	3.29%	0.79%
Spain	**1,993**	2.91%	3.61%	3.26%	0.70%
USA	**14,472**	5.12%	5.75%	5.44%	0.63%
Brazil	**9,434**	3.73%	4.06%	3.89%	0.33%

Turkey	**3,732**	4.00%	3.56%	3.78%	-0.44%
Unknown	**643**	5.41%	4.03%	4.67%	-1.38%
Grand Total	**48,943**	3.92%	4.63%	4.28%	0.71%

The table above presents the number of users, conversion rates (CR), and the differential in conversion rates between the Control (Group A) and Treatment (Group B) for various countries. Out of the 12 countries analysed, 10 countries experienced an increase in conversion rates in the Treatment group compared to the Control group.

Canada exhibited the highest increase of 1.79%, while Turkey showed a decrease of -0.44%. The USA had the largest number of users, accounting for over 25% of consumer activity, and maintained a relatively high conversion rate. These observations suggest that the Treatment strategies had a generally positive impact on conversion rates across multiple countries, with Canada showing the most significant improvement.

Table 2.8: Country Average Spend (Ascending Order):

Country	Number of users	Group A - Control	Group B - Treatment	Average Spend	Differential B-A (+/-)
UK	**2,949**	$2.11	$4.50	$3.32	$2.39
Spain	**1,993**	$2.18	$3.23	$2.71	$1.05
Canada	**1,570**	$3.60	$4.20	$3.91	$0.60
Mexico	**5,738**	$2.81	$3.35	$3.08	$0.54
Australia	**1,168**	$1.67	$2.08	$1.87	$0.41
Unknown	**643**	$3.26	$3.54	$3.41	$0.28
Brazil	**9,434**	$3.21	$3.07	$3.14	-$0.14
USA	**14,472**	$4.30	$4.05	$4.17	-$0.25
France	**3,090**	$2.68	$2.27	$2.47	-$0.41
Germany	**3,854**	$3.40	$2.71	$3.05	-$0.69
Turkey	**3,732**	$3.69	$2.49	$3.08	-$1.20
Grand Total	**48,943**	$3.37	$3.39	$3.38	$0.02

The above table presents the number of users, average spend (AS $), and the differential in average spend between the Control (Group A) and Treatment (Group B) for various countries. Out of the 12 countries analysed, 6 countries experienced an increase in average spend in the Treatment group compared to the Control group, resulting in a net increase of $0.02 across all countries.

The UK exhibited the highest increase of $2.39, indicating a significant positive impact on average spend. On the other hand, Turkey showed the largest decrease in average spend with -$1.20. France and Germany also experienced decreases in average spend, highlighting potential areas for improvement in the Treatment strategies.

Overall, the data suggests that the Treatment group's strategies had a mixed impact on average spend across different countries, with some countries showing notable increases while others experienced decreases.

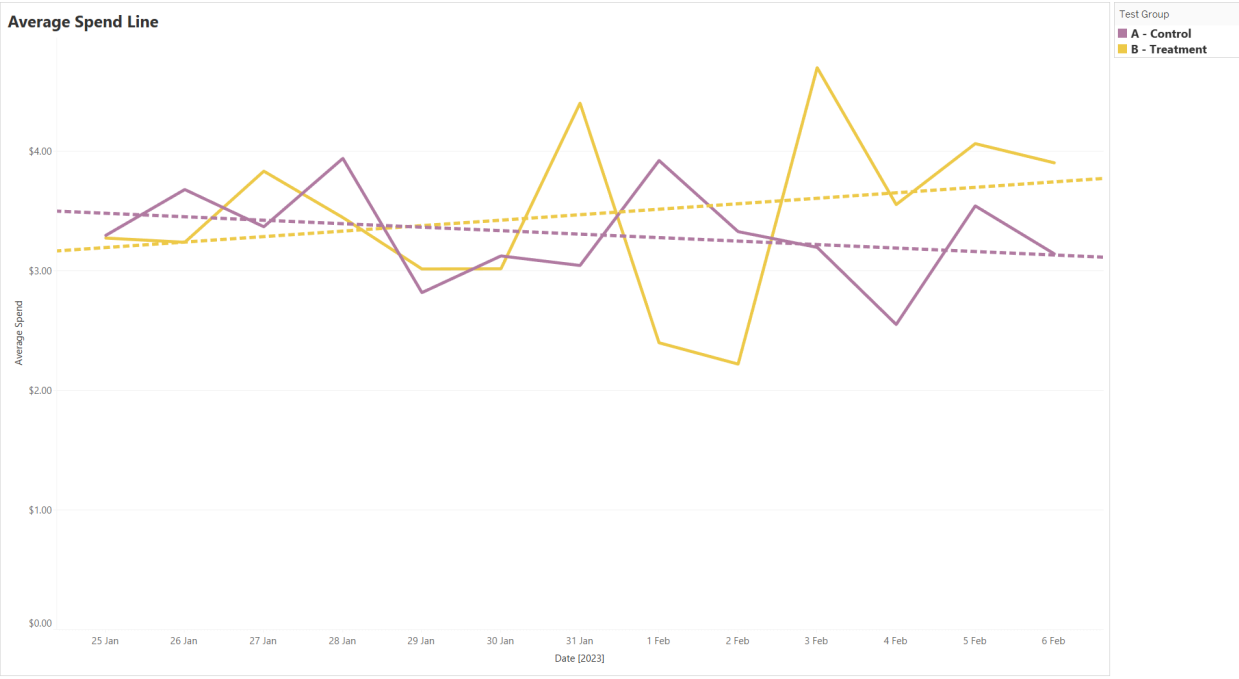
Further analysis can be access via the interactive tableau dashboard found on the following link:

- Globox AB Test Analysis Dashboard

Novelty Effect Analysis

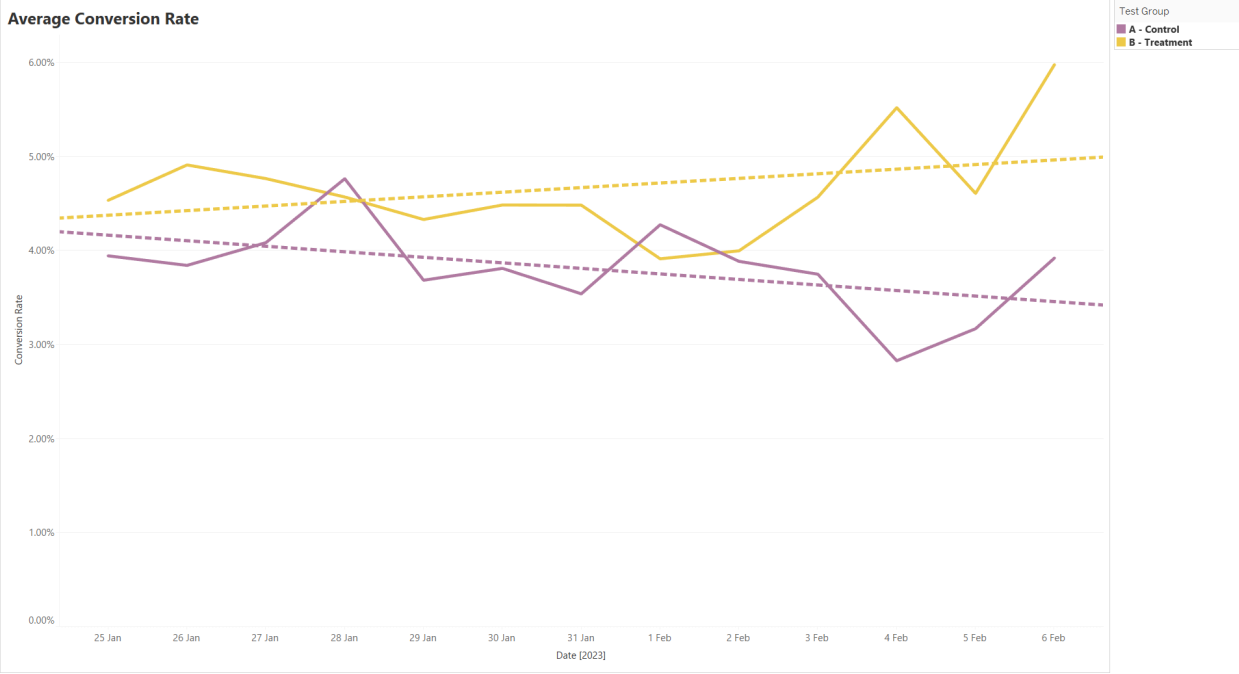
To determine if there is evidence of a Novelty Effect we have analysed 5 metrics Average Amount Spent, Conversion Rate, Average Amount Spent by Converted Users, Cumulative Spend by Converted Users and Time to Purchase to identify if there is

Average Amount Spent:
We began by examining the average amount spent by all users in both groups over time. While some variations were observed, it is crucial to consider overall trends rather than focusing on individual days. Visualising the data through a graph allowed us to understand patterns better. The graph revealed that there was no clear and consistent pattern indicating a significant difference in Average Spend and evidence of the novelty effect on this metric. Although some fluctuations were present, the differences between the Control and Treatment groups were not consistently large enough to suggest a novelty effect.



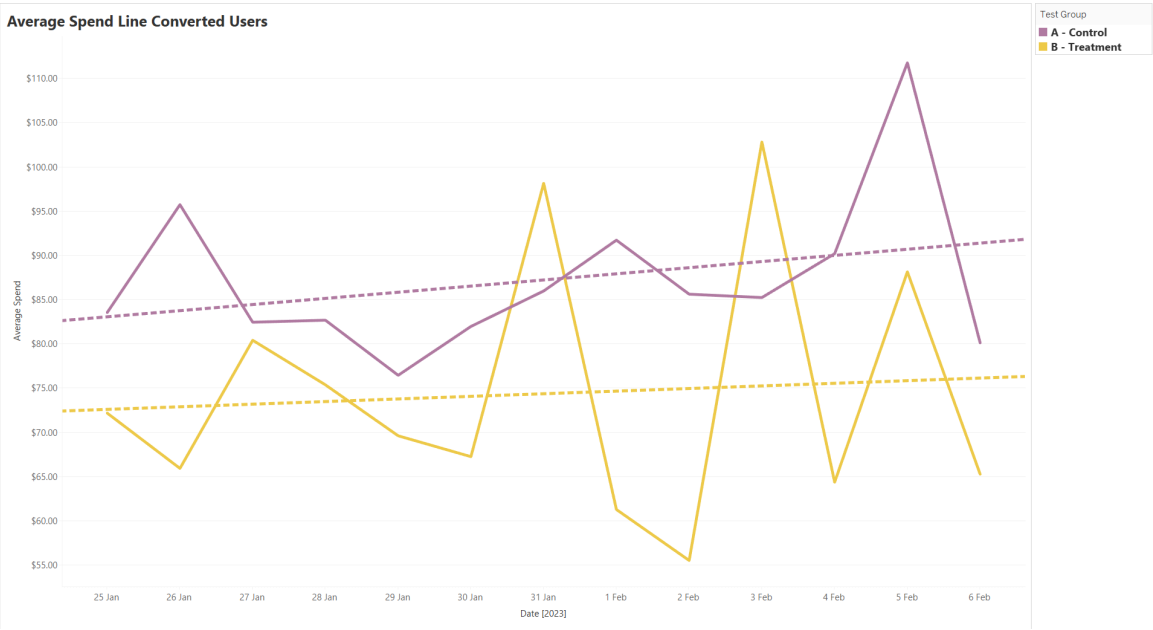
Conversion Rate:

The next metric analysed was the conversion rate for all users in both groups over time. Similar to the average amount spent, the conversion rates did not exhibit consistent and significant differences between the Control and Treatment groups. Although minor fluctuations were observed, there was no indication of a novelty effect significantly affecting the conversion rates.



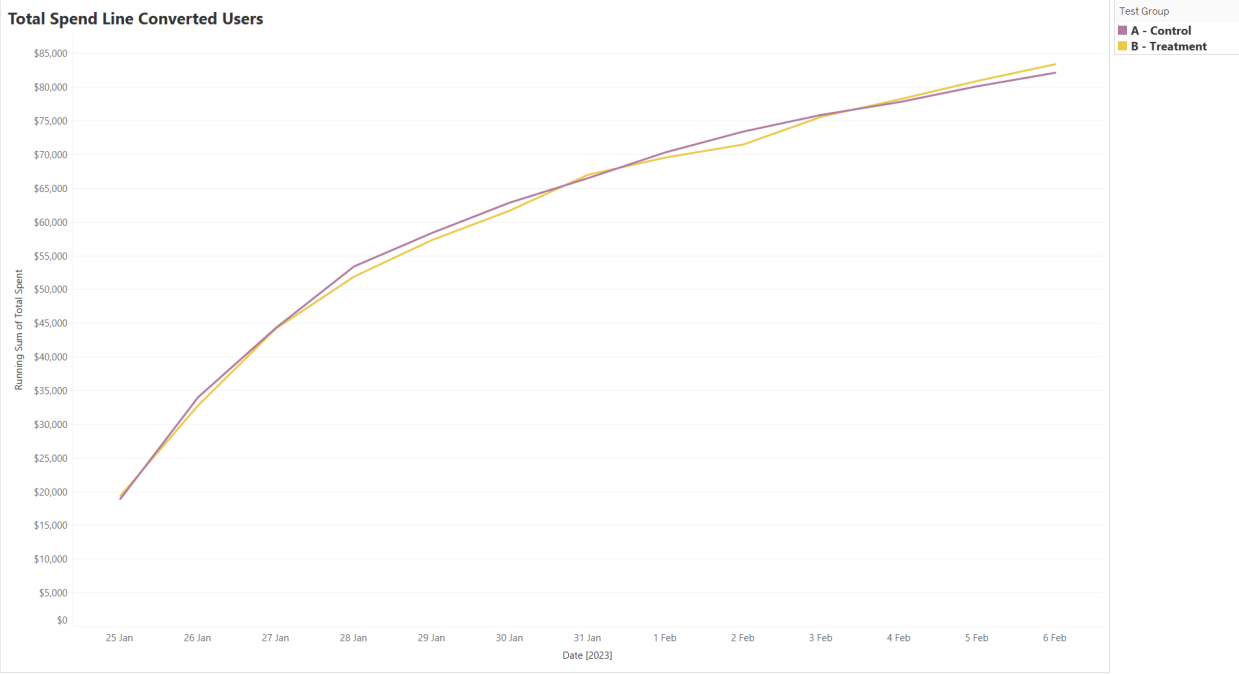
Average Amount Spent by Converted Users:

We then focused on the average amount spent by converted users in both groups. The differences between the Control and Treatment groups in terms of average amount spent by converted users consistently showed the Control group reporting a larger average spend comparative to the Treatment group (tracking a $10 difference). The graph depicting this data also revealed some variations, but again, no pattern emerged to suggest a novelty effect.



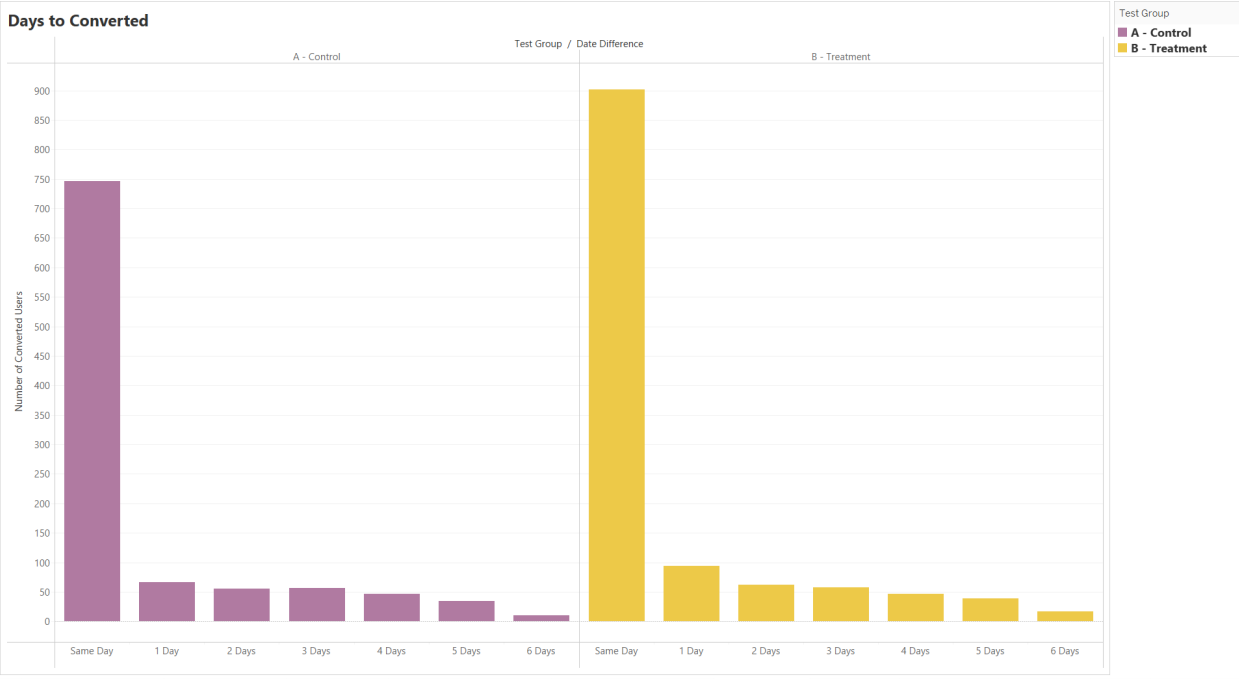
Cumulative Spend by Converted Users:

The cumulative amount spent by converted users between the Control and Treatment groups was also analysed. However, this metric did not highlight any particular differences and showed consistent tracking during the A/B test. The Treatment group reported a difference of $1,269, but it was not substantial enough to indicate a novelty effect.



Time to Purchase:

Lastly, we analysed the time it took for users in both groups to make a purchase after joining the experience. The data did not reveal any substantial differences in the time to purchase between the Control and Treatment groups, with the Treatment group showing an accelerated rate of purchase. Although some variations were observed, such as the number of users purchasing on the same day or within a few days, these differences did not indicate a novelty effect.



Conclusion:
Based on the analysis of the five key metrics, there is no clear evidence to suggest the presence of a novelty effect in the Treatment group compared to the Control group during the A/B test. The average amount spent, conversion rate, average amount spent by converted users, cumulative spend by converted users, and time to purchase did not show consistent and significant differences between the two groups. Therefore, it can be concluded that the novelty being tested did not have a substantial impact on user behaviour during the testing period.

However it is essential to note that the absence of a novelty effect may be influenced by various factors, including the specific context of the test and the characteristics of the user base and the length of the A/B test (13 days). Further research and testing may be required to gain deeper insights into user behaviour and the potential impact of novelty on the product or experience.

Based on the analysis of the data provided, there is no evidence of a novelty effect in the A/B test. The metrics (average spend, conversion rate, total spend of converted users, average amount spent by converted users, and time to purchase), do not consistently show significant differences between the Control and Treatment groups.

Conclusion

Continue iterating

Based on our analysis, it has been determined that the current state of the food and drink banner feature does not exhibit significant improvements in our targeted success metrics, namely the user conversion rate and average spend per user. **The statistical evidence obtained from the A/B test did not provide sufficient support to proceed with the release of the current banner design**.

However, despite the lack of significant improvements, we did observe some promising indications that modifying the banner experience could potentially yield better results in the future. The data revealed trends and patterns suggesting that certain modifications or adjustments to the banner design, placement, or content could have a more positive impact on user engagement and ultimately drive the desired changes in conversion rates and average spend.

These findings highlight the importance of continued iterations and improvements to the food and drink banner feature. It is crucial to explore alternative design elements, content strategies, and user experiences to maximise its effectiveness. By utilising more advanced statistical techniques, such as multivariate analysis or regression analysis, we can gain deeper insights into the specific factors that contribute to user behaviour and identify areas of improvement for the banner feature.

In conclusion, **while the current state of the food and drink banner did not meet our expectations in terms of statistically significant improvements, the observed indications of potential success provide valuable insights for future enhancements**. **By iteratively refining and optimising the banner experience based on these findings**, we can strive to achieve the desired improvements in user conversion rate and average spend, ultimately driving revenue growth for GloBox.

Recommendations

Based on the results of the A/B test, we can conclude **that the new feature had a significant impact on the conversion rate**, with the treatment group 4.63% conversion rate outperforming the control group 3.92%. However, **when considering the average amount spent per user, there was no significant difference between the two groups** ($0.02).

However, we recommend further analysis to understand the impact of the new feature on revenue, such as examining customer behaviour and engagement with the new feature. Based on this additional analysis, stakeholders can make an informed decision about whether to launch the experience to all users

Targeted demographics: Consider segmenting the data based on relevant demographic factors such as location, gender, or device. Analysing the impact of the intervention within specific demographic segments can help identify if the observed trends are consistent across different groups or if the novelty effect varies based on these factors.

Extend parameters: Explore the possibility of extending the parameters of the A/B test, such as increasing the test duration or enlarging the sample size. This allows for a more robust analysis and helps to mitigate potential biases resulting from a limited test period or a smaller sample size.

Consider time of year/seasonality: Account for any seasonal or time-dependent variations that may influence user behaviour. Analysing the results in the context of specific time periods or accounting for seasonal fluctuations can provide a more accurate understanding of the intervention's impact beyond the novelty effect.

Define success and ROI: Clearly define the desired outcomes and metrics that indicate success. Determine the return on investment (ROI) by assessing the impact on key performance indicators (KPIs) such as revenue, customer retention, or customer satisfaction.

Calculate impact on other departments: Consider evaluating the potential impact of the intervention on other departments or areas of the business. Assess if any changes in the conversion rate or average spend could have unintended consequences on aspects like customer support.

Define Minimum Detectable Effect (MDE): Establish the minimum detectable effect, which represents the smallest change in the conversion rate or average spend that is considered meaningful or practically significant. This helps determine the sample size required to detect such a change with a desired level of statistical confidence. See Power Analysis below.

Conduct a post-test analysis: Extend the analysis beyond the A/B test period to evaluate the long-term impact and sustainability of the observed trends. This helps determine if the upward trends in conversion rate and average spend continue or stabilise over time.

Collect qualitative feedback: Gather qualitative feedback from participants to gain insights into their experiences and perceptions of the intervention. This can provide additional context to understand how novelty may have influenced their behaviour and inform further adjustments or interpretations of the results.

Power Analysis/Sample Size Calculation (Conversion Rate)

Introduction:
The purpose of the power analysis is to assess the sample size and statistical power of the test, taking into account the baseline conversion rate, desired Minimum Detectable Effect (MDE), and the observed sample size.

Test sample size: 48,943
Sample Size Conversion Rate: 107,800 (One-sided) 156,500 (Two-sided)
Sample Size Average Size: 369,718

Sample Size and Baseline Conversion Rate:
The A/B test had a total sample size of 48,943 participants. The baseline conversion rate, determined from the control group, was found to be 3.92%. This conversion rate served as the reference point for evaluating the effectiveness of the banner ad.

Minimum Detectable Effect (MDE):
For the analysis, an MDE of 7.5% was chosen. The MDE represents the smallest difference in conversion rates between the experimental and control groups that would be considered statistically significant.

Power Analysis Results:
On a one-sided test: Based on the selected MDE of 7.5% and the baseline conversion rate of 3.92%, **the power analysis indicates that a sample size of 107,800 participants would be necessary to achieve a statistical power of 85%**. However, considering the available sample size of 48,944, it is evident that the test fell short of the desired statistical power.

On a two-sided test: Based on the selected MDE of 7.5% and the baseline conversion rate of 3.92%, an A/B test split ratio of 0.503, a significance level (α) of 0.05 **the power analysis indicates that a sample size of 156,500 participants would be necessary to achieve a statistical power of 85%**. However, considering the available sample size of 48,944, it is evident that the test fell short of the desired statistical power.

To work out sample sizing the following inputs
Baseline: Conversion rate of Control Group 3.92%
MDE: 7.5%
A/B Split Ratio: 0.503
Significance (α): 0.05
Calculator: Statsig

Recommendation:
Given the findings of the power analysis, it is recommended that the A/B test be re-run with a larger sample size. Increasing the sample size will enhance the sensitivity of the test, allowing for the detection of smaller differences between the experimental and control groups.

Conclusion:
In conclusion, the power analysis conducted for the A/B test of Globox's banner ad revealed that **the available sample size of 48,943 participants was insufficient to achieve the desired statistical power of 85%** for a 7.5% MDE. Recommending the re-run of the test with a larger sample size will help ensure that the results are more reliable and conclusive. By obtaining a more substantial sample, Globox will increase the sensitivity of the test, enabling the detection of smaller yet significant differences in conversion rates between the experimental and control groups.

Power Analysis/Sample Size Calculation (Average Spend)

Introduction:
To further define the appropriate sample size by comparing two independent means, utilising the average spend between the control and treatment groups.

Test sample size: 48,943
Sample Size Average Spend: 369,718

Inputs:
For the analysis, a difference of means of 7.5% was chosen, resulting in a value of 0.253. The expected standard deviation used for the power analysis was 25.67 units.

Power Analysis Results:

On a two-sided test: Based on the selected difference of means of 7.5% (0.253), an equal A/B test split ratio, a significance level (α) of 0.05 **the power analysis indicates that a total sample size of 369,718 participants would be necessary to achieve a statistical power of 85%**. However, considering the available sample size of 48,944, it is evident that the test fell short of the desired statistical power.

To work out sample sizing the following inputs
Difference of Means: 7.5%/0.253
Expected Standard Deviation: 25.67
Significance (α): 0.05
Calculator: Statulator

Recommendation:
Given the findings of the power analysis, it is recommended that the A/B test be re-run with a larger sample size. Increasing the sample size will enhance the sensitivity of the test, allowing for the detection of smaller differences between the experimental and control groups.

Conclusion:
In conclusion, the power analysis conducted for the A/B test of Globox's banner ad revealed that the available sample size of **48,943 participants was insufficient to achieve the desired statistical power of 85%**. Recommending the re-run of the test with a larger sample size will help ensure that the results are more reliable and conclusive. By obtaining a more substantial sample, Globox will increase the sensitivity of the test, enabling the detection of smaller yet significant differences in conversion rates between the experimental and control groups.

Appendices

Please find the following list of appendices and links to find source data of information submitted within this report.

- Globox A/B Test Analysis Presentation
    - Presentation Slides - *see pdf Globox - AB Test Analysis - Presentation Slides*
    - Video Presentation Link
- Data Extraction SQL Queries
- Globox A/B Test GoogleSheets Analysis
    - Globox A/B Test Dataset Extraction
    - Hypothesis Testing Calculations - Conversion Rates
    - 95% Confidence Interval - Conversion Rate
    - Hypothesis Testing Calculations - Average Spend
    - 95% Confidence Interval - Average Spend
    - Tables for the report
- Globox A/B Test Converted Users Number of Days

- Globox AB Test Analysis Dashboard
    - Average Spend and Conversion Rate
    - Conversion Rate by Device and Location
    - Novelty Effect Analysis 1/2 - Conversion Rate and Average Spend
    - Novelty Effect Analysis 2/2 - Days to conversion, conversion spend
    - Average Spend Distribution of Converted Users
    - Confidence Intervals - Conversion Rate and Average Spend
    - 95% Confidence Intervals - Conversion Rate and Average Spend
    - 95% Confidence Intervals - Conversion Rate and Average Spend (OLD)

Data Extraction and Validation SQL Queries

Dates of experiment

```
SELECT MIN(join_dt) AS start_date, MAX(join_dt) AS end_date
FROM groups
```

Total Number of users

```
SELECT COUNT(DISTINCT uid) AS total_users
FROM groups
```

Average amount spent per user for the control and treatment groups

```
WITH user_spending AS (
  SELECT g.uid, g.group, COALESCE(SUM(a.spent), 0) AS total_spent
  FROM groups g
  LEFT JOIN activity a ON g.uid = a.uid
  GROUP BY g.uid, g.group
)
SELECT u.group, AVG(u.total_spent) AS average_spent_per_user
FROM user_spending u
GROUP BY u.group;
```

SQL Query to extract aggregated user data with date

```
SELECT u.id AS user_id, u.country, u.gender, COALESCE(a.device, g.device) AS
device_type, g.group AS test_group,
      CASE WHEN SUM(a.spent) > 0 THEN 'Yes' ELSE 'No' END AS converted,
      SUM(a.spent) AS total_spent, g.join_dt AS start_date
FROM users AS u
JOIN groups AS g ON u.id = g.uid
LEFT JOIN activity AS a ON u.id = a.uid AND a.device = g.device
GROUP BY u.id, u.country, u.gender, COALESCE(a.device, g.device), g.group, g.join_dt
```

**Note pre spreadsheet and Tableau analysis updated csv and filled blank cells using Find & Select and filling blank cells with "0"*

SQL Query to work out days converted to users (Date Difference)

```
SELECT n.group AS test_group, COUNT(n.user_id), n.date_difference
FROM(SELECT
        a.uid AS user_id,
    g.group,
    g.join_dt AS date_registered,
    a.dt AS date_converted,
    SUM(COALESCE(a.spent, 0)) AS total_spent_usd,
    a.dt - g.join_dt AS date_difference
     FROM groups AS g
JOIN activity AS a
ON g.uid = a.uid
GROUP BY 1,2,3,4) AS n
GROUP BY 1,3
ORDER BY 3;
```

Data Analysis and Visualisation - GoogleSheets and Tableau

Google Sheets:

- Globox A/B Test GoogleSheets Analysis
- Globox A/B Test Converted Users Number of Days

Tableau Dashboard - Globox AB Test Analysis Dashboard